

Package ‘tmbLab’

March 22, 2021

Type Package

Title Aligning Tumor Mutational Burden (TMB) across diagnostic platforms: A calibration tool developed in conjunction with Phase 2 of the Friends of Cancer Research TMB harmonization project

Version 1.0.0

Description This package generates and applies a calibration model as described in Phase 2 of the Friends of Cancer Research (FOCR) tumor mutational burden (TMB) harmonization project. The software estimates a calibration curve with corresponding prediction limits to quantify the average relationship between WES and panel TMB assay values as well as variability around that curve. Intervals of uncertainty are also provided to accompany the corresponding WES-calibrated TMB values.

Depends R (>= 3.6.0)

Imports nlme

Suggests knitr, devtools, roxygen2, rmarkdown

Author BRP Software Development Team <TMB.Project.Support@mail.nih.gov>

Maintainer BRP Software Development Team <TMB.Project.Support@mail.nih.gov>

License file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

VignetteBuilder knitr

NeedsCompilation no

R topics documented:

tmbLab	2
tmbLabWES	6

Index	10
--------------	-----------

Description

This function estimates, from user-supplied training data, a calibration curve with corresponding prediction limits to quantify the average relationship between WES and panel TMB assay results as well as variability around that curve. The user must also input additional TMB values from the same panel diagnostic assay, and the package will provide estimates of TMB that are calibrated to WES TMB using the calibration curve and associated prediction limits calculated from the training data. Intervals of uncertainty are also provided to accompany the corresponding WES-calibrated TMB values. Refer to the vignettes documentation HTML file for a more detailed description of the package including a quick start guide and images to describe the calibration curve, prediction intervals, and intervals of uncertainty.

Usage

```
tmbLab(  
  file.Model.Data.TMB,  
  file.Obs.Panel.TMB,  
  Calib.Interval = 95,  
  trunc.neg.flag = c("NEG", "TRUNCtoZERO"),  
  dir.output,  
  dir.result = "tmbLab_res",  
  show.HTML = TRUE  
)
```

Arguments

`file.Model.Data.TMB`

This required file describes the laboratory-specific panel TMB training data file name and must be a tab-delimited text file with 3 or more columns: "Sample.ID", the unique sample identifiers; "Uniform.WES.TMB", the training set WES values; "Panel.1", "Panel.2," etc., denoting one or more laboratory-specific panel TMB columns. It is anticipated that most users will provide only one set of laboratory-specific panel TMB data, corresponding to "Panel.1," however, the option is given to include data from multiple laboratory-specific panels by specifying additional columns. Note also that the model implemented in this package assumes that samples are independent. Inputting multiple replicate TMB measurements from the same sample will not lead to appropriate statistical inference (e.g., prediction limits will be incorrect).

`file.Obs.Panel.TMB`

This required file describes the name of the data file that contains laboratory-specific panel TMB values at which the function will provide calibrated WES TMB estimates and intervals of uncertainty. The file must be a tab-delimited text file with 2 columns: "Sample.ID", the unique sample identifiers; "Panel.TMB", panel TMB values for which calibrated WES TMB estimates and intervals of uncertainty will be calculated. Note that "Panel.TMB" values may be real or hypothetical panel TMB values at which WES-calibrated estimates and intervals of uncertainty are desired. Typical hypothetical TMB values of interest would be 5, 10, 15, and 20. The file must contain at least one Panel.TMB value.

Calib.Interval	Specifies the level of uncertainty of the calibrated interval. For example, Calib.Interval = 95 means that 95 will be projected onto the WES TMB axis to derive 95 of uncertainty. The inputted value must be a number between 1 and 100. Common values are 90, 95, and 99. Default is 95.
trunc.neg.flag	<ul style="list-style-type: none"> • If value is "NEG", the code allows negative values of panel-based TMB to be used in the regression modeling. Although TMB values cannot actually be negative, some laboratories apply correction factors to panel TMB values that could cause small values to become negative. Negative values are not a problem for fitting the regression line, but they must be set to zero for purposes of variance and prediction limit calculation. Nonetheless, WES-calibrated estimates of TMB will always be truncated at zero so that they are not reported as negative values. • If "TRUNCtoZERO", the code truncates negative panel-based TMB values to zero prior to their use in regression modeling.
dir.output	This argument can be used to specify the directory output file path. Default is <user's home directory>/tmbLab.
dir.result	This argument can be used to specify the folder name in which the results will be placed. This folder will be found within the folder "dir.output". Default is "tmbLab_res".
show.HTML	If TRUE, the HTML summary document will automatically pop up in the system default browser. Default is TRUE. A copy of this output is also saved for later viewing (see tmbLab.html output file).

Value

The output list includes the following elements:

- Obs.PANEL.TMB.vec: observed panel-based TMB values.
- Obs.PANEL.IDs.vec: observed panel-based TMB sample IDs.
- PANEL.vec: the panel names which were provided in the file file.Model.Data.TMB.
- Calib.Interval: the calibrated interval level of uncertainty.
- trunc.neg.flag: the option for allowing negative values of panel-based TMB.
- dir.resPath: the file path for the final result files.
- dir.output: the file path for the output files.
- predMethod: the calculation method: 1 for "tmbLab"; 2 for "tmbLabWES".
- Lab.All: a data frame including the WES-calibrated estimated TMB with intervals of uncertainty.

The following output files are generated:

- tmbLab.html: The HTML output pop up, which appears automatically if show.HTML=TRUE, includes a summary of relevant information. This same information is saved in a file that will be written to the specified dir.output. The preprocessing section of the output file describes the total number of samples, the number of samples that were dropped because they had WES TMB values >40, and the range of data values observed. The table provides the second data file input values ("Sample.ID" and "Obs.Panel.TMB") along with the WES-calibrated estimates ("Calib.Est.TMB") and the lower and upper intervals of uncertainty corresponding to these values ("CALIB.Lower.Lim.TMB" and "CALIB.Upper.Lim.TMB"). The next column, "Range.Indicator," indicates whether the specified "Obs.Panel.TMB" value falls within the range of laboratory-specific TMB values (e.g., "Panel.1") described in file.Model.Data.TMB.

“Range.Indicator” will be designated “In” if the specified “Obs.Panel. TMB” value falls within the range of laboratory-specific TMB values in the training data (e.g., the range of “Panel.1” described in file.Model.Data.TMB) and will be designated “Out” if the specified “Obs.Panel.TMB” value falls outside this range. For reference, this range is also described in the preprocessing section of the HTML output. The upper and lower limits of uncertainty are set to “NA” or 0, respectively, if the upper limit cannot be interpolated from within the limits of the plot or if the lower limit is below 0. In the last column, a link is provided to a .pdf plot of the scatterplot along with the calibration fitted regression line (solid black line), calibration regression line parameters (text on bottom right), and the prediction limits (dotted black line). Portions of the information that are summarized in this HTML file are also included in various output files in the dir.output folder, “tmbLab_html” and “tmbLab_res,” as described below.

- Panel.n.GLS.ML.scatterplot.pdf: Scatterplots depicting laboratory-specific TMB values (y-axis) versus WES TMB values (x-axis), a 45 degree line (dashed red line), the superimposed regression line (solid black line), and prediction intervals (dotted black line) are in the tmbLab_html folder and are linked to in the final column of the tmbLab.html file. The level of the prediction intervals is specified through the Calib.Interval (Default = 95) level controls the calibration interval level. GLS ML regression parameter estimates are also provided in the bottom righthand corner of the plots, including the intercept, slope, Spearman R, Power variance parameter, and sigma variance parameter. The x and y axes of the plots range from 0 to 55 as these lower values are of greatest relevance for clinical decision making. Users are advised to check for possible outliers that lie beyond the upper range of the y-axis as these outliers have the potential to be influential. The naming convention of the scatterplots is such that n in the name “Panel.n.GLS.ML.scatterplot.pdf” refers to the corresponding panel number specified in the columns of file.Model.Data.TMB. If only “Panel.1” is specified, only one plot will be generated with the name “Panel.1.GLS.ML.scatterplot.pdf.”
- All.GLS.ML.RegistrationParameters.txt: Detailed regression parameters, found in the tmbLab_res and tmbLab_html folder. The model is fit as described in the background section. Each Panel for which data are supplied in file.Model.Data.TMB will correspond to a row in this file. The information and parameters described in this file include N, the number of paired laboratory-specific TMB and WES TMB points included in the training data used for fitting the calibration model; N.NEG.TMB, number of negative panel TMB values provided (for most laboratories this number will be zero); MODEL.AIC, the Akaike information criterion for the fitted regression model; MODEL.BIC, the Bayesian information criterion for the fitted regression model; BETA0, BETA0.SE, BETA0.95CI.LL, BETA0.95CI.UL, which describe the estimated intercept parameter, the corresponding standard error, lower 95 limit, and upper 95 BETA1.95CI.LL, BETA1.95CI.UL, which describe the estimated slope parameter, the corresponding standard error, lower 95 confidence limit, respectively; SPEARMAN.R, which describes the Spearman correlation coefficient estimating the association between panel TMB and WES TMB in the training data; POW.PARAM, POW.SE, POW.95CI.LL, POW.95CI.UL, which describe the estimated power parameter, the corresponding standard error, lower 95 respectively; SIGMA.PARAM, SIGMA.95CI.LL, SIGMA.95CI.UL, which describe the estimated sigma parameter, the corresponding lower 95 and upper 95 which describe the estimated sigma parameter on the natural log (base e) scale and the corresponding standard error, respectively; RESID.VAR, which describe the standard error of the residuals corresponding to the fitted model. Refer to the package nlme() for more information.
- tmbLab.All.GLS.ML.FIT.TMB.txt: File containing points lying on estimated regression lines and prediction limits calculated at the originally supplied WES TMB values. Columns are “Sample.ID”, the panel identifier “Panel.n”, regression estimates “FIT.Est.TMB”, lower prediction limits “FIT.Lower.Lim.TMB”, and upper prediction limits “FIT.Upper.Lim.TMB” calculated at the WES values “Uniform.WES.TMB” provided in training data file.Model.Data.TMB, found in the tmbLab_res folder. If training data from multiple laboratory-specific panels are provided, this file will include regression results for all laboratory-specific panels for

which training data were supplied (“Panel.1”, “Panel.2”, etc.) in a long data format so that each panel’s information is appended with a variable specifying the panel number. Note that this output file provides results similar to those provided by the function `tmbLabWES`, except that `tmbLabWES` allows users to specify WES values for which this information is output, whereas, this file provides results using only the WES values that were input as part of the training data in `file.Model.Data.TMB`.

- `tmbLab.All.GLS.ML.CALIB.txt`: File containing WES-calibrated TMB estimates and intervals of uncertainty calculated at user-supplied laboratory-specific panel TMB values that were input via `file.Obs.Panel.TMB`. Columns are “Sample.ID”, the panel identifier “Panel.n”, panel TMB values at which WES-calibrated values will be calculated (“Obs.Panel.TMB” from `file.Obs.Panel.TMB`), WES-calibrated values “CALIB.Est.TMB”, lower limit of uncertainty “CALIB.Lower.Lim.TMB,” upper limit of uncertainty “CALIB.Upper.Lim.TMB,” and the indicator (“Range.Indicator”) designating whether the specified “Obs.Panel.TMB” value falls within the range of laboratory-specific TMB values supplied in the training data (i.e., in `file.Model.Data.TMB`). If multiple laboratory-specific panels are specified, this file will include regression results for all laboratory-specific panels that were entered (“Panel.1”, “Panel.2”, etc.) in a long data format so that each panel’s information is appended with a variable specifying the panel number.
- `All.GLS.ML.RegressionCurveFit.txt`,
`All.GLS.ML.RegressionCurveFit.Grid.txt`: Background files describing the predicted regression curve estimates calculated at “Uniform.WES.TMB” values supplied for the training data in the `file.Model.Data.TMB` and on a function-generated fine grid (denoted by “.Grid” in the file name) to enable easy plotting of smooth curves. These files are located in the `tmbLab_html` folder. Note that the information in the file “`All.GLS.ML.RegressionCurveFit.txt`” is also summarized for use in the file “`tmbLab.All.GLS.ML.FIT.TMB.txt`” in the `tmbLab_res` folder.
- `All.GLS.ML.LowerLimitPredictionCurve.txt`,
`All.GLS.ML.LowerLimitPredictionCurve.Grid.txt`,
`All.GLS.ML.UpperLimitPredictionCurve.txt`,
`All.GLS.ML.UpperLimitPredictionCurve.Grid.txt`: Background files describing the lower and upper prediction limits for “Uniform.WES.TMB” values supplied for the training data in the `file.Model.Data.TMB` and on a function-generated fine grid (denoted by “.Grid” in the file name) to enable easy plotting of smooth curves. The fine grid will be the same as that used to generate the output file `All.GLS.ML.RegressionCurveFit.Grid.txt`. These files are located in the `tmbLab_html` folder. Note that the information in the files “`All.GLS.ML.LowerLimitPredictionCurve.txt`” and “`All.GLS.ML.UpperLimitPredictionCurve.txt`” is also summarized for use in the file “`tmbLab.Lab.All.GLS.ML.FIT.TMB.txt`” in the `tmbLab_res` folder.

Author(s)

BRP Software Development Team, <TMB.Project.Support@mail.nih.gov>

Examples

```
# Estimate calibrated TMB with intervals of uncertainty
## Not run:
res <- tmbLab(file.Model.Data.TMB =
  file.path(path.package("tmbLab"), "extdata/Model.WES.Panel.TMB.txt"),
  file.Obs.Panel.TMB =
  file.path(path.package("tmbLab"), "extdata/NewSample.Panel.TMB.txt"))

## End(Not run)
```

Description

This function estimates, from user-supplied data, a calibration curve with corresponding prediction limits to quantify the average relationship between WES and panel TMB assay results as well as variability around that curve. For a set of input WES values, this function also provides estimates of laboratory-specific panel diagnostic TMB and prediction limits for the laboratory-specific panel diagnostic assay. This function differs from the all-encompassing function `tmbLab()` in that it estimates the regression line and prediction limits, but does not further provide estimates of laboratory-specific TMB that are calibrated to WES TMB. Refer to the vignettes documentation HTML file for a more detailed description of the package including a quick start guide and images to describe the calibration curve and prediction intervals.

Usage

```
tmbLabWES(  
  file.Model.Data.TMB,  
  file.WES.Panel.TMB,  
  Pred.Interval = 95,  
  trunc.neg.flag = c("NEG", "TRUNCtoZERO"),  
  dir.output,  
  dir.result = "tmbLabWES_res",  
  show.HTML = TRUE  
)
```

Arguments

`file.Model.Data.TMB`

This required file describes the laboratory-specific panel TMB training data file name and must be a tab-delimited text file with 3 or more columns: "Sample.ID", the unique sample identifiers; "Uniform.WES.TMB", the training set WES values; "Panel.1", "Panel.2," etc., denoting one or more laboratory-specific panel TMB columns. It is anticipated that most users will provide only one set of laboratory-specific panel TMB data, corresponding to "Panel.1," however, the option is given to include data from multiple laboratory-specific panels by specifying additional columns. Note also that the model implemented in this package assumes that samples are independent. Inputting multiple replicate TMB measurements from the same sample will not lead to appropriate statistical inference (e.g., prediction limits will be incorrect).

`file.WES.Panel.TMB`

This required file describes the name of the data file that contains WES TMB values at which panel TMB predicted values along with prediction limits will be calculated. The file must be a tab-delimited text file with 2 columns: "Sample.ID", the unique sample identifiers; " WES.TMB ", WES TMB values for which predicted panel TMB and prediction limits will be calculated. Note that "WES.TMB" values may be real or hypothetical WES TMB values at which predicted panel TMB and prediction limits are desired. Typical hypothetical WES values of interest would be 5, 10, 15, and 20. The file must contain at least one WES.TMB value.

Pred.Interval	Specifies the prediction interval level of uncertainty. For example, Pred.Interval = 95 means that 95 TMB for a sample with the designated WES.TMB are predicted to fall within the interval. The inputted value must be a number between 1 and 100. Common values are 90, 95, and 99. Default is 95.
trunc.neg.flag	<ul style="list-style-type: none"> • If value is "NEG", the code allows negative values of panel-based TMB to be used in the regression modeling. Although TMB values cannot actually be negative, some laboratories apply correction factors to panel TMB values that could cause small values to become negative. Negative values are not a problem for fitting the regression line, but they must be set to zero for purposes of variance and prediction limit calculation. Nonetheless, WES-calibrated estimates of TMB will always be truncated at zero so that they are not reported as negative values. • If "TRUNCtoZERO", the code truncates negative panel-based TMB values to zero prior to their use in regression modeling.
dir.output	This argument can be used to specify the directory output file path. Default is <user's home directory>/tmbLab.
dir.result	This argument can be used to specify the folder name in which the results will be placed. This folder will be found within the folder "dir.output". Default is "tmbLab_res".
show.HTML	If TRUE, the HTML summary document will automatically pop up in the system defaultbrowser. Default is TRUE. A copy of this output is also saved for later viewing (see tmbLabWES.html output file).

Value

The output list includes the following elements:

- Obs.PANEL.TMB.vec: observed panel-based TMB values.
- Obs.PANEL.IDs.vec: observed panel-based TMB sample IDs.
- PANEL.vec: the panel names which were provided in the file file.Model.Data.TMB.
- Pred.Interval: the prediction interval level of uncertainty.
- trunc.neg.flag: the option for allowing negative values of panel-based TMB.
- dir.resPath: the path for the final result files.
- dir.output: the path for the output files.
- predMethod: the calculation method: 1 for "tmbLab"; 2 for "tmbLabWES".
- Lab.All: the data frame of the predicted panel TMB with intervals of uncertainty.

The following output files are generated:

- tmbLabWES.html: The HTML output pop up, which appears automatically if show.HTML = TRUE, includes a summary of relevant information. This same information is saved in a file that will be written to the specified dir.output. The preprocessing section describes the total number of samples, the number of samples that were dropped because they had WES TMB values >40, and the range of data values observed. The table provides the second data file input values ("Sample.ID" and "WES.TMB") along with the regression estimates ("Est.TMB") and the lower and upper prediction intervals corresponding to these values ("Lower.Lim.TMB" and "Upper.Lim.TMB"). The next column, "Range.Indicator", indicates whether the specified "WES.TMB" value falls within the range of WES TMB values (e.g., "Panel.1") described in file.Model.Data.TMB. "Range.Indicator" will be designated "In" if the specified "WES.TMB" value falls within the range of WES TMB values in the training data (e.g., the

range of “Panel.1” described in file.Model.Data.TMB) and will be designated “Out” if the specified “WES.TMB” value falls outside this range. For reference, this range is also described in the preprocessing section of the HTML output. The lower prediction interval limits are set to 0 if the lower limit is below 0. In the last column, a link is provided to a .pdf plot of the scatterplot along with the calibration fitted regression line (solid black line), calibration regression line parameters (text in bottom right of plot), and the prediction limits (dotted black line). Portions of the information that are summarized in this HTML file are also included in various output files in the dir.output folder, “tmbLabWES_html” and “tmbLabWES_res,” as described below.

- Panel.n.GLS.ML.scatterplot.pdf: Scatterplots depicting laboratory-specific TMB values (y-axis) versus WES TMB values (x-axis), a 45 degree line (dashed red line), the superimposed regression line (solid black line), and Pred.Interval (Default = 95) (dotted black line) are in the tmbLabWES_html folder and are linked to in the final column of the tmbLabWES.html file. GLS ML regression parameter estimates are also provided in the bottom righthand corner of the plots, including the intercept, slope, Spearman R, Power variance parameter, and sigma variance parameter. The x and y axes of the plots range from 0 to 55 as these lower values are of greatest relevance for clinical decision making. Users are advised to check for possible outliers that lie beyond the upper range of the y-axis as these outliers have the potential to be influential. The naming convention of the scatterplots is such that n in the name “Panel.n.GLS.ML.scatterplot.pdf” refers to the corresponding panel number specified in the columns of file.Model.Data.TMB. If only “Panel.1” is specified, only one plot will be generated with the name “Panel.1.GLS.ML.scatterplot.pdf.”
- All.GLS.ML.RegistrationParameters.txt: Detailed regression parameters, found in the tmbLabWES_res and tmbLabWES_html folder. The model is fit as described in the background section. Each Panel for which data are supplied in file.Model.Data.TMB will correspond to a row in this file. The information and parameters described in this file include N, the number of paired laboratory-specific TMB and WES TMB points included in the training data used for fitting the calibration model; N.NEG.TMB, number of negative panel TMB values provided (for most laboratories this number will be zero); MODEL.AIC, the Akaike information criterion for the fitted regression model; MODEL.BIC, the Bayesian information criterion for the fitted regression model; BETA0, BETA0.SE, BETA0.95CI.LL, BETA0.95CI.UL, which describe the estimated intercept parameter, the corresponding standard error, lower 95 respectively; BETA1, BETA1.SE, BETA1.95CI.LL, BETA1.95CI.UL, which describe the estimated slope parameter, the corresponding standard error, lower 95 SPEARMAN.R, which describes the Spearman correlation coefficient reflecting the association between panel TMB and WES TMB in the training; POW.PARAM, POW.SE, POW.95CI.LL, POW.95CI.UL, which describe the estimated power parameter, the corresponding standard error, lower 95 and upper 95 SIGMA.95CI.UL, which describe the estimated sigma parameter, the corresponding lower 95 95 describe the estimated sigma parameter on the natural log (base e) scale and the corresponding standard error, respectively; RESID.VAR, which describe the standard error of the residuals corresponding to the fitted model. Refer to the package nlme() for more information.
- tmbLab.All.GLS.ML.PRED.txt: Results file containing columns pertaining to “Sample.ID”, the panel identifier “Panel.n”, WES values at which predicted panel TMB values will be estimated (“WES.TMB” from file.WES.Panel.TMB), predicted panel TMB values “Est.TMB”, lower limit of the prediction interval “Lower.Lim.TMB,” upper limit of the prediction interval “Upper.Lim.TMB”, and the indicator (“Range.Indicator”) designating whether the specified “Obs.WES.TMB” value falls within the range of WES TMB values supplied in the training data (i.e., in file.Model.Data.TMB). If multiple laboratory-specific panels are specified, this file will include regression results for all laboratory-specific panels for which data were entered (“Panel.1”, “Panel.2”, etc.) in a long data format so that each panel’s information is appended with a variable specifying the panel number.
- All.GLS.ML.RegistrationCurveFit.txt,

All.GLS.ML.RegressionCurveFit.Grid.txt: Background files describing the predicted regression curve estimates for “Uniform.WES.TMB” values supplied for the training data in the file.Model.Data.TMB and on a function-generated fine grid (denoted by “.Grid” in the file name) to enable easy plotting of smooth curves. These files are located in the tmbLabWES_html folder. Note that the information in the file “All.GLS.ML.RegressionCurveFit.txt” is also summarized for use in the file “tmbLab.All.GLS.ML.PRED.txt” in the tmbLabWES_res folder.

- All.GLS.ML.LowerLimitPredictionCurve.txt,
All.GLS.ML.LowerLimitPredictionCurve.Grid.txt,
All.GLS.ML.UpperLimitPredictionCurve.txt,
All.GLS.ML.UpperLimitPredictionCurve.Grid.txt: Background files describing the lower and upper prediction limits for “Uniform.WES.TMB” values supplied for the training data in the file.Model.Data.TMB and on a function-generated fine grid (denoted by “.Grid” in the file name) to enable easy plotting of smooth curves. The fine grid will be the same as that used to generate the output file All.GLS.ML.RegressionCurveFit.Grid.txt. These files are located in the tmbLabWES_html folder. Note that the information in the files “All.GLS.ML.LowerLimitPredictionCurve.txt” and “All.GLS.ML.UpperLimitPredictionCurve.txt” is also summarized for use in the file “tmbLab.Lab.All.GLS.ML.PRED.txt” in the tmbLabWES_res folder.

Author(s)

BRP Software Development Team, <TMB.Project.Support@mail.nih.gov>

Examples

```
# Estimate Panel TMB with intervals of uncertainty
## Not run:
resWES <- tmbLabWES(file.Model.Data.TMB =
  file.path(path.package("tmbLab"), "extdata/Model.WES.Panel.TMB.txt"),
  file.WES.Panel.TMB =
  file.path(path.package("tmbLab"), "extdata/NewSample.WES.TMB.txt"))

## End(Not run)
```

Index

tmbLab, [2](#)
tmbLabWES, [6](#)